

Application No. 10/732,809Docket No. 324-163**REMARKS**

Applicants thank the Examiner for the consideration given the present application. Claims 1-9 are pending, of which claims 1 and 8 are Independent. Minor changes are made to claims 1, 6, and 7 for clarity. Claim 8 is a method claim corresponding to independent apparatus claim 1. Claim 9 is directed to a computer arrangement for automatically performing the steps of claim 8.

Reconsideration is requested of the rejections under 35 U.S.C. §103(a) of claims 1-3 and 6 as being unpatentable over Van der Akker (U.S. 6,415,250) in view of Walton (U.S. 5,392,419) and claims 4, 5, and 7 under 35 U.S.C. §103(a) as being unpatentable over Van der Akker and Walton in view of De Campos (U.S. 6,272,456).

Van der Akker discloses an automatic language identification system 110 (FIG. 3) based on a probability analysis of predetermined word portions extracted from an input text 301, the language of which is to be identified. A word portion is the ending of a word having a predetermined number of characters (column 11, line 65, through column 12, line 7; column 20, lines 4-9), generally a suffix or, in the beginning of a word, a prefix (column 8, line 45, through column 9, line 3; FIG. 2C).

A language corpus analyzer 302 associates each word portion of a predetermined language corpus 309 with a normalized frequency indicative of the number of times the word portion is found with the corpus (column 9, lines 35-42, and column 12, lines 44-60) and with a relative probability derived from the frequency in relation to the size of the

Application No. 10/732,809Docket No. 324-163

corpus. When the word portion rarely appears in the language corpus, the probability is close to zero (column 10, lines 56-60; column 13, lines 38-41).

A language Identification engine 306 sums the relative probability values associated with each language for each of the word portions extracted from input text 301 and found in probability table 304. Engine 306 identifies the language of input text 301 having the largest total accumulated relatively likelihood value (column 10, lines 33-45). As a result, the Identification language system 110 relates to one category of first character strings (suffixes or prefixes) in a word (column 20, lines 52, 53, 66, and 67; claim 1).

Analyzer 302 of Van der Akker analyzes only one character string per extracted word with respect to a corpus 309, whereas the analyzing means in the system described in Applicant's claim 1 analyzes plural character strings for an extracted word. Analyzer 302 of Van der Akker thus fails to carry out the function of Applicant's claimed analyzing means.

Moreover, system 110 of Van der Akker applies to each character string a probability depending on its frequency in a language corpus, not the location in the word extracted from the input text. The character string location in the extracted word is particularly useful when all the character strings included in the extracted word are compared to first and second character strings. However, as paragraph [0016] of Applicant's published application indicates, the present invention is not limited to trigrams or word portions having a particular location.

Application No. 10/732,809Docket No. 324-163

The atypical character strings in Van der Akker are associated with a probability value of zero or a common negative value, -0.99. Applicant's claim 1 requires a score associated with the determined language to be increased by a first coefficient depending on the position of the first character string found in the extracted word, and, whenever a second character string is found in the extracted word, the score to be decreased by a respective second coefficient that is associated with the found second character string and that increases as the probability of the found second character string in the determined language decreases. Hence, the second coefficient is different for each of the atypical character strings, resulting in significantly improved language identification accuracy. The language identification system of Van der Akker is far less accurate than that of the present application, the object of which is to improve language identification accuracy.

Therefore, Van der Akker fails to disclose analyzing means and comparing means of Applicant's Independent claim 1. Furthermore, Van der Akker does not disclose means for storing "first" frequently character strings and means for storing "second" atypical character strings before analyzing and comparing the strings.

The Office Action admits Van der Akker does not disclose that the second coefficient increases the probability of the character strings being in the language decreases, but says the atypical character strings in Van der Akker are associated with a probability value of zero or a common negative value, -0.99. Walton is relied on for this feature. Walton merely describes a language identification system for a data block

Application No. 10/732,809Docket No. 324-163

included in an incoming command/data stream. The languages are printed in such languages as PCL language or PostScript language (column 3, lines 1-11). For each print language, the Walton system particularly analyzes the presence of defined portions, each referred by a "for" key and promoting the identification of the language, and the presence of other defined portions, each referred by an "against" key and depressing the identification of the language. For each print language and at each detection of a "for" key or an "against" key in the data block, a "for" tally register, respectively, an "against" tally register, sums a value associated with the detected key. This value is multiplied, respectively divided, by a skew value indicating the importance of the key in the context of the data block (column 2, lines 20-24, and column 4, lines 1-11). At the end of the data block analysis, for each print language, the "for" and "against" tally registers both associated with each of the languages are compared to the other registers to identify the best language used in the data block (column 2, lines 25-35).

The Office Action compares the skew value of Walton with the second coefficient of Applicant's claim 1. The second coefficient associated with a second character string found in the input text does not change each time the second character string is found in the input text. The second coefficient has a fixed value corresponding to the improbability of finding the second character string in the input text according to a defined language. The context of the input text does not interact with the second coefficient.

De Campos fails to cure the deficiencies of the other references.

RECEIVED
CENTRAL FAX CENTER

JUL 27 2007

Application No. 10/732,809Docket No. 324-163

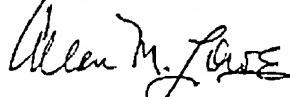
Accordingly, Applicant's language identification system is new and unobvious from Van der Akker and Walton, neither of which suggests the analyzing and comparing means and steps of Applicant's independent claims 1 or 8.

In view of the foregoing remarks, reconsideration and withdrawal of the rejections are respectfully requested.

To the extent necessary, a petition for an extension of time under 37 C.F.R. §1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including application processing, extension, and extra claims fees, to Deposit Account 07-1337, and please credit any excess fees to said deposit account.

Respectfully submitted,

LOWE HAUPTMAN & BERNER, LLP



Allan M. Lowe, Registration No. 19,641

1700 Diagonal Road, Suite 300
Alexandria, VA 22314
(703) 684-1111 telephone
(703) 518-5499 telecopier
AML:pcf